Lauren Brodsky
Final Project Report
BAIS: 3250
5/9/25

Introduction: This project aims to explore the relationship between weather conditions and Citi Bike usage in New York City. I was inspired by this topic when my brother went on a trip to New York City. During his time exploring the city, he used a Citi Bike and he was surprised to see that even in the cold and rainy weather it was still hard to find an open bike. This made me curious about how much weather really affects bike sharing behavior. The goal of this analysis is to understand how factors like rain, temperature, and wind impact Citi Bike usage. These insights can help companies plan for better bike availability and continue encouraging bike sharing as a more reliable mode of transportation.

Data: This project used two datasets. The first one is the NYC Citi Bike Ride Share System Data (2023) from Kaggle. It has records for individual Citi Bike trips throughout the year 2023 in New York City. Each data entry includes information such as ride ID, bike type (classic, electric, or docked), timestamps for ride start and end times, station names and IDs, start and end coordinates, and user type (member or casual). The second dataset used is NYC Weather Data (2023), which was scraped from Weather Underground. The scraped data contains information about daily average temperature, wind speed, and precipitation. After collecting the data, I horizontally merged these two datasets on “date” and named the final dataset as “final_merged_data.csv”. By combining them, I was able to then analyze how different weather conditions impact Citi Bike usage across New York City.

Final Merged Data Dictionary:

Field	Type	Description
date	DateTime	The date the weather and Citi Bike usage was recorded
total_rides	Numeric	Total number of bike rides on that date
classic_bikes	Numeric	Number of rides using classic bikes
electric_bikes	Numeric	Number of rides using electric bikes
docked_bikes	Numeric	Number of rides using docked bikes
average_ride_duration	Numeric	Average duration of rides (in minutes)
rides_by_members	Numeric	Number of rides taken by members
rides_by_casual	Numeric	Number of rides taken by casual users
avg_temperature	Numeric	Average temperature for the day (in degrees Fahrenheit)
precipitation_in	Numeric	Precipitation recorded for the day (in inches)
wind_speed_mph	Numeric	Average wind speed recorded for the day (in miles per hour)

Analysis:

1) Impact on Temperature:

Research Question 1a): How does daily usage change with temperature?

Results: To answer this question, I used Linear Regression, K-Nearest Neighbors (KNN), Ridge Regression, and Lasso Regression. I also created a monthly trendline using Matplotlib to visualize the relationship between average temperature and total rides by month. The results showed that as temperature increases, so do daily ride counts. Linear Regression had an R^2 of 0.5115 and KNN had an R^2 of 0.4099, meaning Linear Regression explained the changes in ride counts better than KNN. I also tested Ridge and Lasso Regression to see if the results would improve using these methods. Ridge had an R^2 of 0.5116 and Lasso had an R^2 of 0.5115, which were basically the same as the Linear Regression. Since all models had nearly the same results, adjusting the model didn't make a big difference for this analysis. Additionally, the regression coefficient was 47.37 which suggests that for every 1°F increase in temperature, there are about 47 more rides. To further visualize this trend, I created a trendline which shows the average temperature and total rides by month.

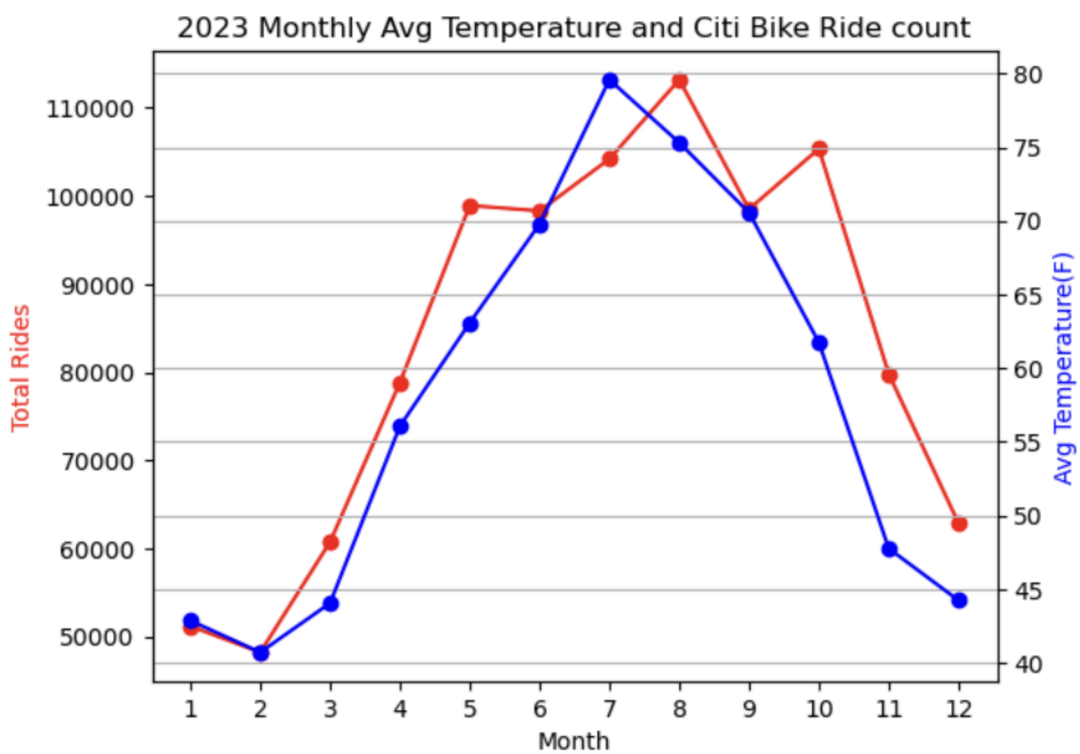


Figure 1: 2023 Monthly Average Temperature and Citi Bike Ride Count

As seen in *Figure 1*, the monthly trendline supports these findings by showing peaks in both temperature and ride counts during the warmer months and declines in colder months. These results match what I expected, where people ride less in the colder months due to the uncomfortable temperature.

Research Question 1b): <u>Is there a noticeable trend in rides peaking at certain temperature ranges?</u>

Results: To answer this question, I created a scatterplot of daily ride counts vs. average temperature using individual data points. Unlike question 1a, which uses monthly averages to show overall trends, this analysis looks at individual daily data points to show more detailed patterns of how temperature affects ride counts.

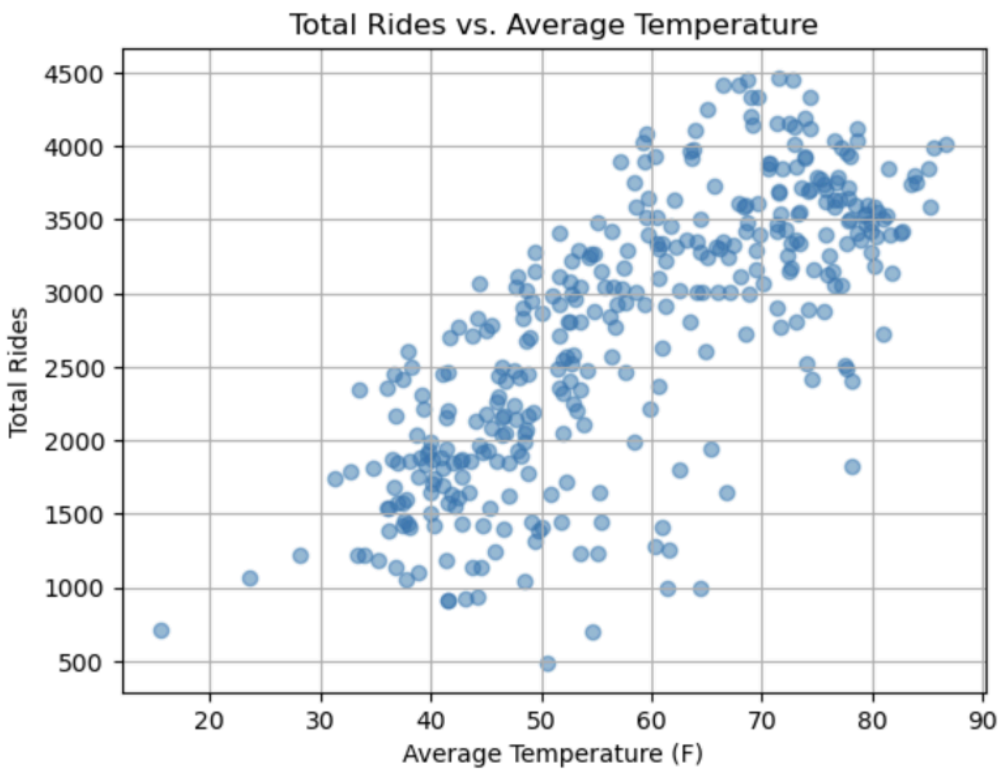
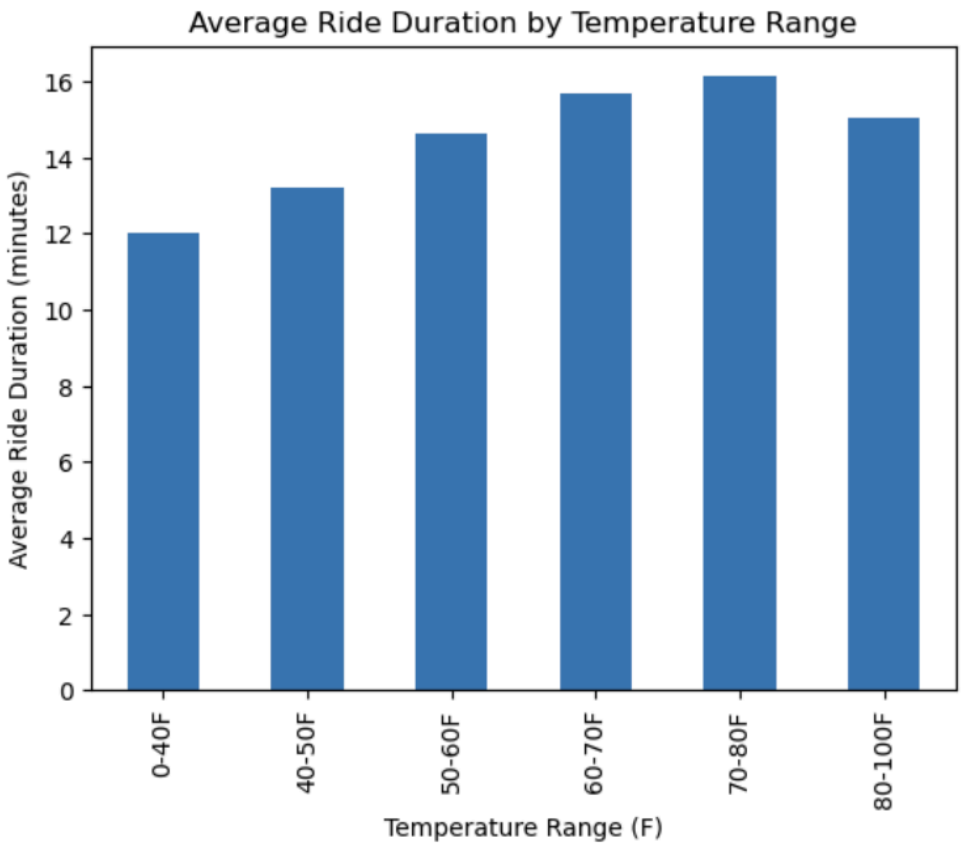


Figure 2: Total Rides vs. Average Temperature

Based on *Figure 2*, the results showed a steady increase in rides as temperatures rose, especially between 60°F and 85°F. This matches my expected result that people ride more in comfortable weather and less during really hot or really cold temperatures.

Research Question 1c): Does ride duration change with temperature?

Results: To explore this question, I used summary statistics, a binned bar chart, and a t-test. The summary statistics showed that the average ride duration was about 14.5 minutes and most rides ranged between 9 and 25 minutes.




```
temp_range
0-40F      11.997752
40-50F     13.220502
50-60F     14.613766
60-70F     15.667624
70-80F     16.118619
80-100F    15.019543
```

Figure 3: Average Ride Duration by Temperature Range

As seen in *Figure 3*, the binned bar chart showed that ride duration increased with temperature and peaked at 70–80°F with an average of 16.12 minutes, which is an 11% increase from the overall average. This trend suggests that more comfortable weather conditions could mean longer ride times. Next, a t-test was run to compare cold (0–40°F) and warm (70–80°F) days to see if the difference in average ride duration between the two temperature ranges was statistically significant. The results showed a

T-value of -12.37 and a p-value of 0.0000. Therefore, we can conclude that the difference is statistically significant. The negative t-value showed that ride duration in colder weather is significantly lower. Overall, people ride about 4 minutes longer on warm days (70–80°F) than they do on cold days (0–40°F), showing a link between warmer temperatures and longer Citi Bike usage.

2) Impact on Precipitation:

Research Question 2a): What is the impact of rain on bike usage?

Results: To explore the impact of rain on bike usage, I created a scatter plot and calculated the correlation between daily precipitation and total ride counts.

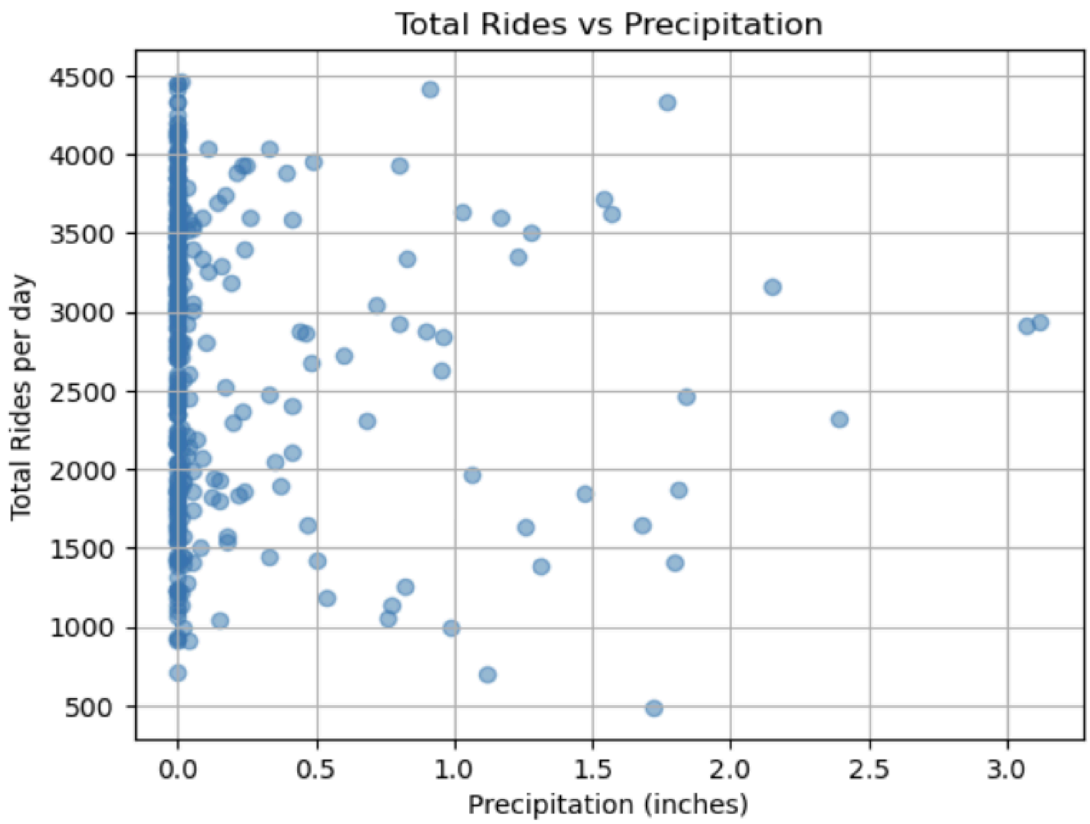


Figure 4: Total Rides vs. Precipitation

The scatter plot in *Figure 4* showed that most rides occur on days with little or no rain, and ride counts tend to drop as rainfall increases. This trend suggests that as the rain amount increases, the number of Citi Bike rides tends to decrease. To better understand the relationship, I calculated the correlation coefficient, which was -0.0788. This indicates a weak negative relationship between rain and ride counts, meaning that rain lowers bike usage. People are less likely to ride on rainy days, although rides still occur despite the weather.

Research Question 2b): Are classic bikes more popular than electric bikes during rainy weather?

Results: To determine whether classic bikes are more popular than electric bikes during rainy weather, I created a grouped bar chart and calculated the percent change in bike usage from no rain to heavy rain.

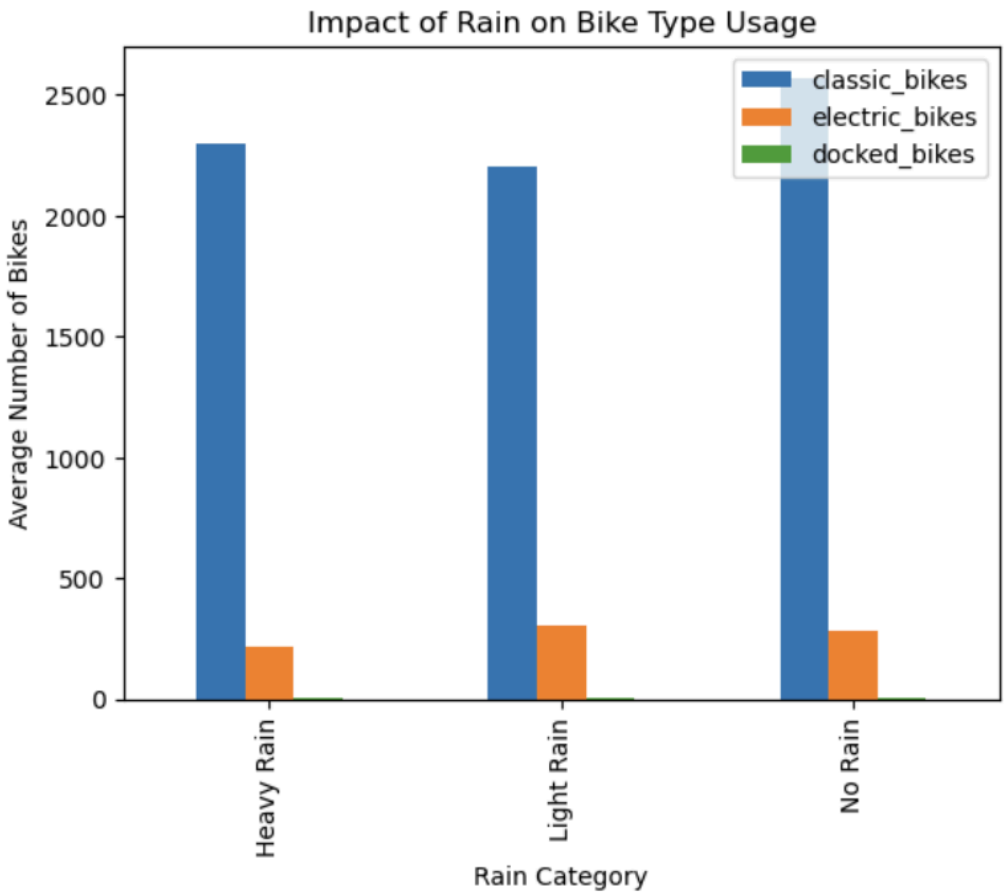


Figure 5: Impact of Rain on Bike Usage Type

As seen in *Figure 5*, the chart showed that classic bikes had the highest usage in all weather conditions, followed by electric bikes. It also showed that docked bikes were used the least overall, which is why they are barely visible on the chart. The bar chart showed that bike usage decreases as rain increases, since all bike types showed their highest usage on days with no rain. Next, the percent drop analysis supported these findings, showing that classic bikes had the smallest decrease in usage during heavy rain (-10.7%), compared to electric bikes (-21.7%) and docked bikes (-38.2%). This means that classic bikes are more popular overall and are also more consistently used in rainy conditions.

Research Question 2c): Do people ride longer when it rains?

Results: To answer if people ride longer when it rains, I used grouped means and an ANOVA test. I grouped the data by rain amount and calculated the average ride duration. This showed that ride durations were slightly longer during heavy rain (14.69 minutes) compared to light rain (14.23 minutes) and no rain (14.57 minutes). My initial hypothesis was that ride durations would decrease on rainy days and there would be a more significant drop with heavy rain compared to light rain. To test if this difference was statistically significant, I ran an ANOVA test. The results showed an F-statistic of 0.6429 and a p-value of 0.5264. Since the p-value is greater than 0.05, we fail to reject the null hypothesis. This means there is no statistically significant difference in ride duration across rain levels. Even though ride durations appear slightly longer in heavy rain, the difference is not statistically meaningful. Therefore, we can reasonably conclude that rain does not have a statistically significant impact on ride duration.

3) User Types:

Research Question 3a): Are casual users (day pass riders) more sensitive to weather than members?

Answer: To explore this question, I used both grouped means and percent change analysis for rain levels, temperature ranges, and wind speeds. I first calculated grouped means to understand the average number of rides by user type under different weather conditions. Then, I calculated the percent change to fairly compare how much usage drops or increases relative to the baseline (such as no rain, low temperature, or low wind). I did this because casual and member users have different overall ride volumes, and I wanted the data to reflect those differences more accurately.

The results showed that casual riders tend to respond more strongly to weather changes overall. For rain, casual rides dropped more in light rain (-16.45%) than member rides (-11.04%), compared to the baseline of no rain. However, members dropped more in heavy rain (-12.57%) than casual users (-8.58%). For temperature, casual rides peaked with a 297.35% ride increase from 0–40°F to 70–80°F, while member rides only increased by 80.27%. Lastly, the wind data showed that casual rides dropped by 46.99% at wind speeds of 30+ mph, while member rides dropped by 32.34%, compared to the baseline of low wind (0–10 mph). These patterns suggest that casual users are more sensitive to weather, and members show more consistent riding behavior overall. These results matched what I expected. Since members pay for a monthly membership, they are more likely to ride regardless of the weather. Whereas, casual riders pay per day, so they might be more influenced by weather conditions when deciding whether to ride.

4) Forecasting Future Usage:

Research Question 4a): <u>How many Citi Bike rides are expected in the next 90 days based on past trends?</u>

Answer: To answer this question, I used two time series forecasting models, ARIMA and TBATS. I used total ride counts to predict the number of Citi Bike rides that will occur in the next 90 days. The ARIMA model predicted that approximately 162,196 rides will occur during the next 90 days. The ARIMA Mean Absolute Error (MAE) was 1,049, meaning that on average, the daily predictions were off by about 1,049 rides. Since the MAE was high, I also applied a TBATS model to see if I could get more accurate results. The TBATS model forecasted approximately 286,593 rides over the same 90 day period. The TBATS MAE was lower at 789, which means it is more accurate than the ARIMA model. Since the TBATS model performed better, I chose to create a graph to visualize the predicted trends in ride counts.

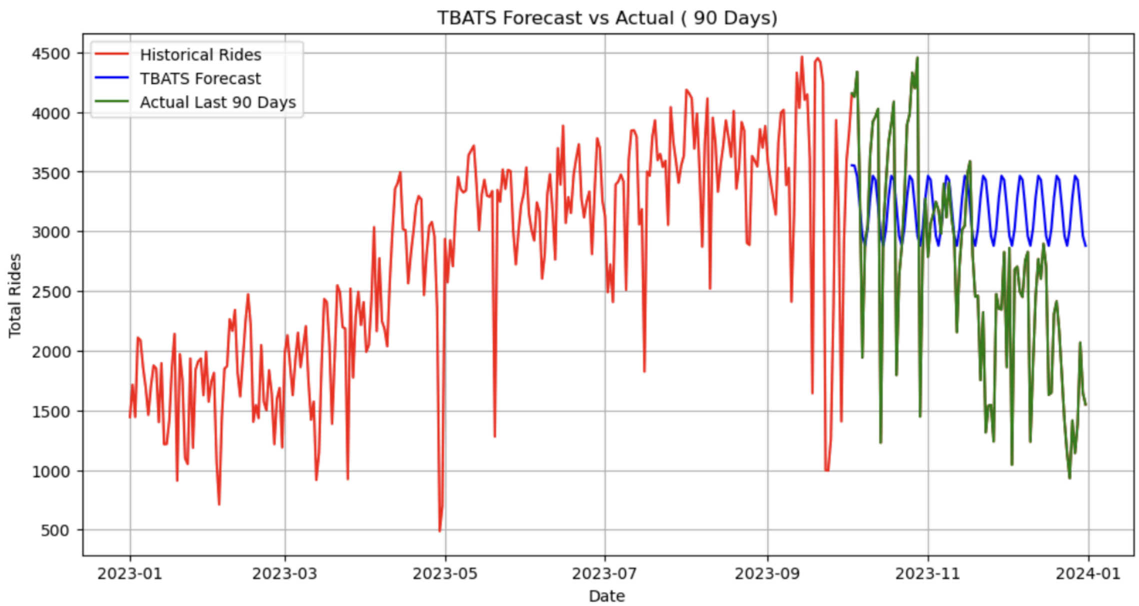


Figure 6: TBATS Forecast vs. Actual (90 Days)

As shown in *Figure 6*, the TBATS forecast (blue line) follows a consistent pattern, while the actual ride counts over the last 90 days (green line) show more of a fluctuation and an overall decrease in rides toward the end of the year. The historical data (red line) showed an increase in rides throughout the warmer months (mid-year) before dropping off toward the colder months. Overall, this prediction can be used to help Citi Bike allocate resources more efficiently. Knowing how many rides to expect can also help with planning bike placement, repairs, and staffing, especially during busy or changing weather seasons.

Conclusion:

This project explored how weather conditions impact Citi Bike usage in New York City using ride and weather data from 2023. The results showed that warmer temperatures lead to higher ride counts and longer ride durations. It also found that rain, high wind speeds and cold temperature tend to reduce bike usage. Casual users were more sensitive to weather changes than members, especially in response to temperature and strong wind. Next, classic bikes remained the most consistently used bike type across all weather conditions. Overall, the final analysis confirms that weather plays an important role in understanding bike share usage.

Limitations:

A limitation is that there is a smaller number of records for extreme conditions like very hot, cold, windy, or rainy days. This may impact how accurate or representative the averages are. The dataset also does not account for outside factors such as holidays, weekends, or local events. These factors could influence ride patterns independently of weather. In addition, the data does not show whether people are riding to commute, exercising, or for leisure. The different ride motivations could also affect how riders respond to weather, especially for casual users.

Future Work:

Future analysis could benefit from using hourly weather data instead of daily averages. This would help capture more precise weather conditions and could improve the overall accuracy of the results. Next, exploring the usage patterns at each bike station could also reveal how local factors such as location and population density influence bike usage behavior. Finally, expanding the analysis to include more recent data or additional cities could also help determine whether these trends apply across other bike share systems or cities.

References

citibikenyc. (2023). *NYC Citi Bike Ride Share System Data 2023*. Kaggle. https://www.kaggle.com/datasets/hassanabsar/nyc-citi-bike-ride-share-system-data-2023?resource=download

New York City, NY Weather History. (2023). Weather Underground. https://www.wunderground.com/history/daily/us/ny/new-york-city/KLGA/date/2023-1-1